THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking

Citation for published version:
Gunson, N, Marshall, D, Morton, H & Jack, M 2011, 'User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking', *Computers and Security*, vol. 30, no. 4, pp. 208-220. https://doi.org/10.1016/j.cose.2010.12.001

Digital Object Identifier (DOI):
10.1016/j.cose.2010.12.001

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Computers and Security

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 07. May. 2023

User Perceptions of Security and Usability of Single-Factor and Two-Factor Authentication in Automated Telephone Banking

Nancie Gunson[a,*], Diarmid Marshall[a], Hazel Morton[a], Mervyn Jack[a].

[a] Centre for Communication Interface Research, The University of Edinburgh, Alexander Graham Bell Building, King's Buildings, Edinburgh, EH9 3JL, UK.

* Corresponding author. Tel.: +44-131-651-7120; fax: +44-131-650-2784; E-mail address: Nancie.Gunson@ccir.ed.ac.uk (N. Gunson).

Keywords: Authentication; Two-factor; Security; Automated telephony; Usability; Empirical study; Dialogue design.

Abstract

This paper describes an experiment to investigate user perceptions of the usability and security of single-factor and two-factor authentication methods in automated telephone banking. In a controlled experiment with 62 banking customers a knowledge-based, single-factor authentication procedure, based on those commonly used in the financial services industry, was compared with a two-factor approach where in addition to the knowledge-based step, a one-time passcode was generated using a hardware security token. Results were gathered on the usability and perceived security of the two methods described, together with call completion rates and call durations for the two methods. Significant differences were found between the two methods, with the two-factor version being perceived as offering higher levels of security than the single-factor authentication version; however, this gain was offset by significantly lower perceptions of usability, and lower ratings for convenience and ease of use for the two-factor version. In addition, the two-factor authentication version took longer for participants to complete. This research provides valuable empirical evidence of the trade-off between security and usability in automated systems.

1 Introduction

This paper describes an experiment to investigate user perceptions of the usability and security of single-factor and two-factor authentication methods in automated telephone banking.

Most financial institutions worldwide offer some form of telephone banking to their customers, providing them with remote access to their accounts via the telephone from any location. The majority of these services are at least partly automated, which means callers hear a series of automated messages (typically, human speech recorded in a studio by a professional voice talent) and are invited to respond using either their voice or the keys on their telephone keypad, depending on the service design. An increasing number of these services moreover, are fully automated - allowing customers to carry out simple banking tasks such as balance enquiries and funds transfers over the phone without the need to speak to a human agent, often 24 hours a day, 7 days a week.

While the use of Internet banking is growing rapidly[1], automated telephony systems of this type continue to play an important role in banks' service offerings. As an example, the real-life service deployed by one of the UK's major banks (and on which the application in this research is based) has 4 million registered users and handles 5.5 million calls per month, with continued development of the service being the focus of ongoing research at the Bank.

Access to the existing service is via knowledge-based authentication ("what you know"), specifically in this case recall of a secret number or 'PIN'. Other banks employ similar techniques involving the use of a PIN or alphanumeric password or some combination of the two, with knowledge-based authentication currently being the *de facto* standard for customer authentication in telephone banking and other remote channels across the U.K. financial services industry.

However, whilst passwords and PINs remain a potent tool in the security of automated services when they are employed properly (O'Gorman, 2003), for users who have to remember multiple passwords and PINs across a number of applications the cognitive load of remembering each can become disadvantageous. Adams and Sasse (1999) suggest that four or five is the maximum number of unrelated, regularly used passwords that users can reasonably be expected to cope with. In practice, however, some studies suggest that in the workplace alone the number of passwords required can be as high as sixteen (Sasse et al., 2001).

A common solution amongst users is to write the information down or use the same password or PIN for a number of applications, both of which have obvious security risks (Adams and Sasse, 1999; Dhamija and Perrig, 2000; Gaw and Felten, 2006). For example it is reported (Adams and Sasse, 1999) that at least 50% of respondents in a survey wrote down passwords as a result of password expiration policies in the workplace and the consequences of difficulties in remembering multiple passwords. Another study (Gaw and Felten, 2006) found that the majority of respondents had committed three or fewer passwords to memory for online use, and passwords were typically reused twice: the

[1] A Gartner survey showed that 73 percent of U.S. respondents regularly logged on to banking accounts and 63 percent paid bills online (consumeraffairs.com, 2005)

level of password reuse, moreover, rose over time as users acquired more 'accounts' but did not create more passwords. In fact, these figures seem conservative when compared to those of other studies (Dhamija and Perrig, 2000) which have found that while respondents had ten to fifty instances where passwords were required, in reality each had only one to seven unique passwords that were reused repeatedly. Other studies (Florêncio and Herley, 2007) report even more extreme figures where in a study of half a million Internet users each was found on average to have 25 passwords, reusing each for an average of 6.5 different sites.

Users are also known to choose memorable (and therefore low security) passwords (Adams and Sasse, 1999; Bishop, 2005; Yan et al*.,* 2004). PINs, moreover, have been shown to be more difficult to remember than passwords (Sasse et al*.,* 2001) and anecdotal evidence exists of customers frequently using their date of birth as their PIN or writing down their PIN.

Research on how to address some of these issues has included attempts to improve users' choice of passwords through stricter selection policies, better education and/or suggested improvements during the selection process (Forget et al., 2007; 2008; Shay et al., 2010). However, awareness that such techniques can produce only limited results (Kuo et al., 2006; Forget et al., 2008), together with recent growth in remote fraud (Hiltgen et al., 2006) has meant that security concerns continue to grow across the various remote channels (Clarke et al., 2009; Sinclair and Smith, 2005). Users themselves have indicated concerns, with one study of 23,000 European Internet users reporting that 40% are deterred from using online banking by security worries (silicon.com, 2005).

Increasingly, therefore, alternative and/or supplementary security mechanisms are being sought. One alternative to traditional passwords and PINs that has received considerable attention in recent years is *graphical* passwords (Brostoff and Sasse, 2000; Dhamija and Perrig, 2000; Dunphy et al., 2008, 2010; Wiedenbeck et al., 2005). These, unlike alphanumeric passwords and codes, cannot be written down or shared easily. There is also some evidence to suggest that they are more memorable than their text-based equivalents (Chiasson et al., 2009). As a relatively new idea, however, there are few examples of graphical passwords actually in use (Dunphy et al., 2010) and as a consequence little is currently known about the potential for cross-interference and/or other memorability issues similar to those experienced with alphanumeric passwords should their use become commonplace in everyday life. They are also, of course, visual in nature making them unsuitable for use in the automated telephony context considered in this present paper.

Another area of interest is *multifactor* authentication, which involves the use of more than one *type* of authenticator. Authentication components can generally be classified as belonging to one of three categories (the first of which, knowledge-based authentication, has already been discussed in some detail):

i) *what you know* – information secret to the individual (e.g. password)

ii) *what you have* – a physical token unique to the individual (e.g. ATM card)

iii) *what you are* – some intrinsic property of the individual (e.g. fingerprint)

Single-factor authentication uses one of these components as a means of verifying customer identity. The classic example is a PIN or password, as discussed above, although the use of biometric technology for single-factor authentication has also been investigated (Toledano et al., 2006).

Two-factor authentication potentially provides enhanced security by combining two different authentication components; the advantage being that security can be maintained by one of the components if the other component is compromised (O'Gorman, 2003)[2]. A common example of multifactor authentication, found in the U.K., is the bankcard equipped with 'Chip and PIN' whereby cash withdrawal from an ATM or purchase of goods in person requires the user both to be in possession of the card and to know the secret PIN associated with it. Although some weaknesses have been identified in the technology (Murdoch et al., 2010), introduction of this system has had a considerable impact, with a significant decline in the incidence of bankcard related frauds in the six months following the system being made compulsory (Mannan and van Oorschot, 2007). Online banking fraud, on the other hand, rose by 55% in the same period (presumably in part as a consequence of the Chip and PIN initiative).

Concerns over increased fraud via remote channels have thus led the banking industry to consider the use of two-factor authentication in their remote channels. Recent U.S. banking guidelines have suggested that single-factor authentication for Internet banking can be inadequate for certain transactions (FFIEC, 2005) whilst in the U.K. there are already several examples of banks that have added use of a card plus reader token to existing Internet banking security procedures for certain transactions. As yet, however, there do not appear to be any examples of two-factor authentication in a telephone banking context.

In adopting new security procedures care needs to be taken since the changes may impact the perceived usability of the service. This is important since the usability of security measures is key for customer acceptance (O'Gorman, 2003); if the security processes are difficult to use, customers will avoid them or will fail to use them properly[3]. Significantly also, it has been suggested that poor usability can lead to a reduction in use of the remote channel (Knight, 2008).

Usable security, however, is seen as something of an oxymoron. A commonly-held view is that usability and security are competing goals (Kainda et al., 2010; Sasse, 2004), with improvements in one dimension inevitably leading to a negative impact on the other. As recently as 2010, for example, a study of password policies across the Internet (Florêncio and Herley, 2010) employed a single measure (password strength) as a gauge of both security and usability, the assumption being that a secure password policy means poor usability and vice versa. The basis for this simplification, however, is somewhat limited. While it is true that a body of work on authentication using text-based passwords, several examples of which are mentioned above, has demonstrated that more secure passwords

---

[2] Generally, multifactor authentication that combines all three factors has not been widely applied (O'Gorman, 2003).

[3] A study found that although a password policy that allows weak passwords can lead to system compromise, an overly strong policy on the other hand can lead users to write passwords down - thereby inadvertently increasing system vulnerability overall (Shay and Bertino, 2009 - from Shay 2010).

are more difficult for users to remember, these studies have however, tended to focus on specific measures such as password memorability and/or user behaviour in selecting passwords, rather than usability of the system or service as a whole. In fact, the field of what has been termed 'Human Computer Interaction and Security' or 'HCISec' (Kainda et al., 2010) is relatively recent; with commentators noting that very little work was focusing on the usability of secure systems (Kainda et al., 2010; Sasse, 2004). There are several examples of papers discussing the relationship between usability and security (Cranor and Garfinkel, 2005); however, there are relatively few studies available that provide *empirical* data on the topic. One example (Ion et al., 2010) compared the usability of different methods for secure device pairing, with the measure of usability in this case participants' explicit choices between methods in different contexts. Another study (Piazzalunga et al., 2005) used an experimental approach to compare three different "security devices" (a card reader and two types of USB token) for email protection. They based their evaluation metrics on the usability sub-characteristics of the ISO 9126 standard (ISO 1991) for the evaluation of software products (*learnability*, *operability* and *attractiveness*); the study itself, however, involved only ten users.

A comprehensive evaluation of the usability of different security mechanisms (Weir et al., 2009, 2010) reported results from two different studies of Internet banking. Usability evaluation in each case was based on the ISO (1998) definition of usability incorporating *effectiveness*, *efficiency* and *satisfaction* and was assessed using a range of subjective and objective measures, together with user perceptions of security, convenience and ease of use. In the first study (Weir et al., 2009) the usability and security of two-factor authentication involving three different devices was examined, where a password was the first factor in each case. The results showed that in this context the usability of the system *was* sacrificed when increasing levels of security were applied. In the second study (Weir et al., 2010) the single-factor authentication procedure used in the existing service (two passwords) was compared with two different forms of two-factor authentication; password plus digital token, and password plus SMS text message on a mobile phone. Here, interestingly, the single-factor version was rated significantly *less* usable than both forms of two-factor authentication, contradicting the commonly-held assumption that increased security leads to poorer usability.

The focus of most work on user authentication however has been on Internet services. With the exception of O'Gorman (2005, 2006a, 2006b), little or no research on usability and security has taken place in a voice communication context (in five years of the SOUPS[4] conference, for example, not a single paper has focused on the telephony channel). This is important since there are considerable differences between the two channels; they have different modes of input, the telephony channel involves a time pressure to respond that does not exist in most Internet applications, it has no visual aids to memory; and its users, moreover, may be less technologically sophisticated than their Internet counterparts.

This paper seeks to address some of the gaps identified in the literature on usability and security by describing an empirical evaluation of two different methods for user authentication within the context of a real-world automated telephone banking service.

[4] Symposium on Usable Privacy and Security

The usability and security of the existing single-factor, knowledge-based procedure was compared to a two-factor approach involving use of a hardware token. A range of metrics was employed in order to provide a comprehensive evaluation, and in contrast to much of the other work described above (where participants are typically University students) participants in the experiment were real-life customers of the UK bank involved in the study.

2 Experiment Approach

Usability engineering emphasises the importance of directly observing actual and potential users (Karat, 1988). As such, a large component of any usability engineering work involves observing user behaviour in the workplace, at home and in usability laboratories. The experiment approach used in this research involves a contrastive study where two versions of the dialogue system, differing in some design characteristic, are experienced by participants in a laboratory setting. Participants are given detailed personal data as fictitious personae to use during the experiment and are asked to perform tasks typical of real-life use within the dialogue system. The results obtained from this procedure are considered to approximate the responses the service would generate in a real world context of use.

In such controlled usability experiments, a repeated-measures design may be employed in which all participants try the different versions of a design being tested, the benefit being to allow comparisons to be made for each participant (Landauer, 1988). The order that participants experience designs can influence results - with habituation or fatigue effects causing bias. As a result, order of experience is balanced across the cohort in this type of design (Preece et al., 2002; Robson, 1983). Researcher bias can also be a problem in an experimental setting, thus procedures are standardised: each participant receives minimal instruction (priming) and follows the same session blueprint. In this way, the data collected can be used for statistical comparisons (Whiteside et al., 1988; Coolican, 1990). A rich set of data is collected based on performance measurements (such as time taken to complete tasks and success rates) and subjective attitudes to the experiences of using the different versions of the service. During the experiment, researchers make direct observations about the behaviour of the participants; these provide valuable insights into the non-verbal reactions of participants using the interactive systems.

Participants' attitudes are measured using questionnaires completed after experiencing each version of the service. The questionnaire employs a Likert format (Likert, 1932) where each usability attribute to be measured is presented to the participant in the form of a stimulus statement followed by an agree-disagree scale. The advantages of this format have been described (Coolican, 1990) as:

- Participants prefer the Likert scaling technique because it is "more natural" to complete and because it maintains their direct involvement in the process.
- The Likert technique has been shown to have a high degree of validity and reliability.
- The Likert scale has been shown to be effective in measuring changes over time.

Usability of a system is the efficiency, effectiveness and satisfaction of the system to perform specific goals in a particular environment (ISO, 1998). Previous research (e.g. Foster et al., 1993; Love, 1997; Love et al., 1992, 1994) has identified salient attributes of

the perceived usability of interactive systems, and a usability questionnaire in Likert format has been constructed to measure these attributes. The questionnaire covers cognitive issues (e.g. level of concentration required by users, and how stressful the service was to use), the fluency and transparency of the system (e.g. ease of use and degree of complication), system performance (e.g. the efficiency of the application and users' preferences for a human agent), and issues relating to the voice of the service (e.g. politeness and clarity). For the experiment detailed in this paper, the questionnaire was extended with two additional items, one referring specifically to the security of the system and the other relating to the amount of information that users were required to input into the system. See Appendix A for a full listing of the questionnaire.

The 7-point Likert scales are used with a balance of positively and negatively worded stimulus statements in the questionnaire. On this scale, once the responses are normalised for statement polarity, a score over 4.0 represents a positive attitude; scores below 4.0 represent negative attitudes to the identified attributes, and each participant's overall attitude to the service can be measured by taking the mean of these numbers across all of the items in the questionnaire. A measure of the overall attitude to the service can then be obtained by averaging all the questionnaire results for participants who experienced that service. When replies to subjective questionnaire attributes are gathered from multiple users in this way, the average results can be considered an objective measure of system appeal (Nielsen, 1993). Moreover, when scores are collected for several competing versions of a system design, these can be compared and used to determine which is most satisfying to use.

3 Experiment Details

3.1 Authentication Approaches Compared

Each of the authentication strategies investigated was set within the context of an already established automated telephone banking service from a major UK bank. The two versions of the service employed in the experiment were based on the existing customer-facing service and differed only in the customer authentication part of the dialogue. Both began with a welcome message, followed by capture of the caller's account number and sort code (for identification purposes). Following entry of a valid account number and sort code, the authentication dialogue took place.

Two approaches to customer authentication were compared in the experiment. The single-factor approach is based on a "*what you know*" technique deployed in the existing service - recall of two digits, selected at random, from a six-digit secret number previously registered by the user and known only to them ("*Please give the Xth digit of your secret number*." followed in a separate stage by "*...and the Yth digit*.").

The two-factor approach examined in the experiment contains an additional "*what you have*" component as well as the two digits from a secret number. Various hardware authentication tokens are currently available for use in enhanced security. Table 1 details three types of hardware tokens, together with pros and cons for each.

Device	Method of code generation	Pros	Cons
Portable key fob	User presses button on device and access code is displayed	Inexpensive	Hardware to keep / carry around
Smartcard and reader	User inserts card into reader, types in PIN and access code is displayed	Compatible with a Chip and PIN world	Hardware to keep A more expensive solution
Mobile phone	User requests a text message containing access code	Inexpensive	Relies on ownership of a mobile phone. Signal reception.

Table 1: Hardware Authentication Tokens

In this research, the portable device was used as the authentication token - a security device which outputs a one-time access code. The device is a small keyfob-style device weighing 10g, including the battery. It features an (up to) 8-character LCD display and a single push button. An on-board real time clock provides time synchronous encryption. After one push on the button a unique one-time six-digit access code is shown on the display. In the system investigated in the experiment, the user inputs (all of) this one-time access code when prompted by the telephone banking service, as an additional security step after the account number, sort code and secret number prompts. The security key for the generation of an access code on a customer's access device is linked via their account to the same security key for code generation on the bank's server. The authenticating server uses the same algorithm as the device, using the time and serial code of each customer's device to verify their inputted access code. The code is updated every 30-60 seconds. Typically the server accepts only the last three codes. The customer cannot lock the access device, as no PIN entry is required.

Using the single-factor approach, since there are ten possible values for each of the two digits (0-9), the chance of a fraudster guessing both digits correctly on a single attempt is 1 in 100. With three attempts (as allowed by the service before the customer account is 'locked' and the caller is transferred to a human agent for further authentication) the risk rises to 3 in 100 or 3%.

With the two-factor strategy, assuming the fraudster is not in possession of the device, the possibility of guessing a six-digit passcode correctly on any given attempt is 1 in 1 million. With three attempts at a single code allowed this rises to 3 in 1 million. Further, when combined with the secret number stages described above, the chance of a fraudster successfully passing both authentication stages is multiplicative i.e. 3/100 x 3/1,000,000 = 9 in 100 million or 0.000009%.

Numerically, thus, the two-factor approach is significantly more secure than the single-factor strategy, by several orders of magnitude. This, although it does not take into account the possibility of the token being lost or stolen, is obviously of considerable attraction to the bank, leading them to consider deploying use of the token in telephone banking. It is, however, recognised that the additional security offered by such an

approach is only of benefit if customers are both able to use the token and understand its value; hence the motivation for this research.

After successful authentication, as in the existing system the customer can select from various banking services such as checking their balance or searching recent transactions applied to their account.

3.2 Research Aims

The experiment was designed to explore user perceptions of the usability and security of two different authentication approaches in automated telephone banking.

The experiment hypotheses considered here were:

H1: The two authentication methods will be different in terms of perceived usability.

H2: The two authentication methods will be different in terms of a set of four comparative ratings (quality, convenience, security and ease of use).

H3: The two authentication methods will be different in terms of user preference.

Each hypothesis was tested using statistical tests on the measures of usability, quality and preference taken for each authentication method after direct experience.

3.3 Participants

Sixty-two telephone banking users took part in the experiment, in a design that was approximately balanced for age and gender. All were customers of the UK bank involved in the study. Equal numbers of participants from each gender were recruited, 31 males, 31 females. The sample consisted of slightly more older participants than younger participants; 61.3% were age 35 years or over, 38.7% were under age 35 years.

Participants were asked how often, if at all, they telephoned the Bank. The vast majority (85.5%) had experience of phoning the existing, automated service. A majority (54.8%) called a few times a year or less, but a substantial minority (30.6%) called at least once a month.

3.4 Procedure

All participants made one telephone call to each version of the service in a repeated-measures design. A persona description sheet included a note of 'their' secret number; and a bank card was included with their (dummy) account details (showing their account number and sort code). A different persona with different account details was supplied for each version of the service. A task sheet was provided for the participants prior to making each call.

Prior to experiencing the two-factor version, participants were also given the keyfob device with the following (spoken) description:

"For this version of the service you will also need to use this device, which provides you with a unique access code to enter during your call. The access code changes frequently, making it very difficult for anyone else to know all your details. Have a look at the device and try it out."

In order to focus on the identification and verification part of the telephone call, participants were asked to carry out a simple banking task (finding out a balance on their account) in each call.

In order to quantify the differences between the two methods, attitude questionnaires were posed after exposure to each of the different authentication designs. A short debriefing interview after the main experiment elicited information from participants about their reactions to the different authentication methods; their likes, dislikes and suggestions for improvements. Participants were also asked which option they preferred, by rating the two methods in terms of overall quality, convenience, ease of use and security on a 0-30 point scale labelled 'worst' to 'best', with the results analysed as a quality rating and as a rank order of preference.

4 Results

4.1 Caller Behaviour and Performance

During the sessions, the phone calls were automatically logged to report caller behaviour and performance metrics.

4.1.1 Timing Data for Caller Identification and Verification (ID&V)

Participants were timed in their calls from the start to completion of the identification and verification process (ID&V). In the single-factor version, this comprised account number, sort code and 2-from-6 secret number digit input. In the two-factor version, this involved the account number, sort code and secret number steps plus an additional access code input of 6 digits.

The overall mean completion timing for the single-factor version was 41.96 seconds, and for the two-factor was 60.06 seconds. The timings for each participant for each version were compared in a repeated-measures ANOVA, with the authentication version as the within-subject variable, and with age, gender and order of experiencing the two versions as the between-subject factors. The results showed that the difference between versions was highly significant ($p < 0.001$). However, this is as expected since the two-factor approach to authentication does not replace the single-factor approach but instead adds an additional stage to it. Therefore the durations for identification and verification of the two-factor version would be expected to be longer. A significant interaction of version with order of experiencing the two versions was also found ($p < 0.001$). When experienced first, the single-factor version took an average of 44 seconds; when experienced second the mean time for this version was just less than 40 seconds. The two-factor version experienced first took an average of 64 seconds; experienced second it took 56 seconds. The order effect indicates a learning experience, with both versions taking longer when experienced first than when experienced second. Given that the two-factor approach incorporates the single-factor procedure this is not entirely surprising; it does, however, illustrate the importance of balancing the order in which participants experience the two different versions in order to obtain a true gauge of the mean authentication duration in each case.

An interaction of gender with version was less strong, but statistically significant at $p = 0.039$. Female participants took the shortest time with the single-factor version, less than

their male counterparts. Both male and female participants took similar lengths of time to complete the two-factor version (no significant difference).

4.1.2 Behavioural Data

Using the single-factor version, 90% of participants completed the task in their first attempt at the call. The other 10% failed due to problems at the account number and sort code input stages: they were all able to complete the task on the second call attempt. For the two-factor version, 74% of participants completed the task on their first attempt, 18% on the second call attempt and another 8% required a third attempt to complete the task. Seven out of the sixteen participants that had failed call attempts in this version failed at the account number and sort code input stages, that is, due to problems not associated with the authentication method. The other nine participants failed at the secret number or access code input stages (seven of these were in the older age group). The secret number stage requests two random digits from a 6-digit string. This was confused here with the access code, also a 6-digit string. The dynamic access code would therefore need to be better differentiated from the secret number to avoid these errors.

For the two-factor version, participants' first responses at the access code input stage were observed, that is, the first time they encountered this stage regardless of any previously failed call attempts (for example, errors at the account number, sort code or secret number stages). Some 90% of participants completed this stage at the first attempt. The remaining 10% of participants failed this stage at the first attempt and required further attempts; three of them gave no response within the 8-second response window, and the other three made errors inputting the code. Five of these six participants were successful in further attempts at this stage within the same call, the final participant being the only one to require a separate, additional call to successfully navigate the access code stage.

4.2 Usability Results

4.2.1 Overall Usability Scores

The mean usability scores derived from the usability questionnaires were 5.56 (on a 7-point scale) for the single-factor version and 5.31 for the two-factor version. A repeated measures ANOVA was carried out on the scores for each version, taking age, gender and order of experience as between-subject factors. A significant difference on usability between the two versions was found ($p = 0.024$). The usability of the single-factor authentication method was judged to be significantly higher than the two-factor access device version.

In addition, the interaction between the usability of each service and the order of experience indicated a moderately significant effect ($p = 0.036$). Participants in both order groups rated the single-factor version higher than the two-factor version; however, the difference was considerably larger amongst those who used the two-factor version first (single factor 5.70. two-factor 5.26) than in the group who used the standard, single-factor version first, where scores for the two versions were fairly similar (single factor 5.40, two-factor 5.37).

Paired samples t-tests on this data indicated no significant differences between scores awarded to the two versions in the order single-factor followed by two-factor. There was,

however, a highly significant difference in the scores awarded to the versions in the order two-factor followed by single-factor, with the single-factor version scoring significantly higher than the two-factor version, $p = 0.002$. This appears to be a contrastive scoring effect, as these participants had the opportunity to contrast the single-factor security procedure favourably with the two-factor procedure they had already experienced.

The interaction between the authentication method and the age group of participants showed another significant effect ($p = 0.013$), as illustrated in Table 2.

Age Group	Single-factor	2-factor
18-34 years	5.46	5.48
35 years and over	5.58	5.16

Table 2: Mean Usability Scores for Each Version by Age Group

Younger participants scored both interfaces similarly, with the two-factor version achieving a very slightly elevated mean score (not significant). However, older participants scored the two-factor method significantly lower than their younger counterparts, at the same time giving the single-factor version higher scores than younger users. Paired samples t-tests showed that there was a highly significant difference in scores for the two versions given by older participants, awarding significantly higher scores to the single-factor version. A majority of participants who appeared to confuse the secret number and access code of the two-factor version were in the older age group (as described in the previous Behavioural Data section). Although based on small numbers of participants, this may in part explain the differences between the two age groups with respect to the overall usability data.

4.2.2 Analysis of Individual Usability Attributes

A repeated measures ANOVA was carried out on each of the 22 attributes measured by the Likert usability questionnaire for the two authentication methods tested in the experiment, again taking age, gender and order of experience as the between-subject factors.

A number of differences between the two versions were found. The single-factor version scored better in cognitive terms, with less *concentration* required ($p = 0.002$), less *confusion* ($p = 0.007$), less *complication* ($p = 0.011$) and less *frustration* with usage ($p = 0.023$) than the two-factor version. The single-factor version was also rated more positively in terms of not being *too fast* ($p = 0.029$), it was found to be *easier to use* ($p = 0.025$) with fewer *details to input* ($p < 0.001$) than the two-factor version. However, *security* was not as highly perceived as with the two-factor version of the service. The two-factor version scored significantly higher in terms of *security* ($p < 0.001$) at 6.07 on the 7-point scale compared with 5.59 for the single-factor version.

Investigation was made on how the two age groups reacted to the different authentication methods. For the single-factor version, older participants tended to give slightly higher scores than younger participants, significantly higher for *knew what to do* ($p = 0.037$) and *friendliness* ($p = 0.050$); however they scored significantly lower than younger participants in terms of *stress* ($p = 0.046$), that is they felt significantly more stressed than

younger participants when using the service. For the two-factor version, older participants scored significantly lower than younger participants in terms of *confusion* ($p = 0.032$), *flustered* ($p = 0.011$), *stress* ($p = 0.026$) and *frustration* ($p = 0.039$). Both age groups gave very positive reactions to the *voice clarity*, *politeness* and *security* of the two-factor service.

The two versions were then compared within each age group separately. The younger participants perceived the two-factor version as requiring more *concentration* ($p = 0.043$) and having too many *details to input* ($p = 0.038$) compared to the single-factor version. However, they perceived the two-factor version as more *efficient* ($p = 0.050$) and *secure* ($p = 0.008$). Amongst the older participants, many significant differences were found between the two versions; in this group the single-factor version was perceived as being more usable in terms of *confusion* ($p < 0.001$), *concentration* ($p = 0.018$), *flustered* ($p = 0.027$), *frustration* ($p = 0.015$), *complication* ($p = 0.003$), *knew what to do* ($p = 0.002$), *ease of use* ($p = 0.020$), *willingness to use again* ($p = 0.011$), *friendly* ($p = 0.017$), *enjoyment* ($p = 0.020$) and amount of *details to input* ($p = 0.003$). One exception was that older participants felt the two-factor version was more *secure* ($p = 0.032$) than the single-factor version.

4.3 Quality Rating Results

Overall quality ratings as well as ratings for convenience, security and ease of use for the two versions of authentication were recorded on a 30-point scale. The mean results are shown in Table 3. (These data are based on a cohort of 61 participants, as one participant was unable to complete the debriefing interview due to time constraints.)

Rating	Single-factor	2-factor
Overall Quality	21.66	20.00
Convenience	23.89	19.59
Security	22.29	25.31
Ease of Use	25.42	22.78

Table 3: Ratings for the Two Authentication Methods

In the same way as the usability results, a repeated measures ANOVA was carried out on each of these quality ratings with age, gender and order of experience as the between-subjects factors.

In terms of *overall quality*, there was no significant difference between the two versions and no significant interactions or between-subjects effects were found.

In terms of *convenience*, there was a highly significant difference between the two authentication methods ($p < 0.001$), with the single-factor version being seen as significantly more convenient than the two-factor version. There were no significant interactions or between-subjects effects. Participants were asked to comment on their reasons for their response. One participant stated that the two-factor version "*was too much effort. The [single-factor version] is less hassle*".

In terms of *security*, there was also a highly significant difference between the two authentication methods ($p < 0.001$), this time with the two-factor version being seen as significantly more secure than the single-factor version. Again participants were asked to comment on their reasons for their response. One participant stated that "*[the single-factor] security wasn't too bad but when I get to [two-factor version] it gives you that extra bit of security*."

There was also a significant interaction between perceived *security* and age group ($p = 0.013$), the interaction is shown in Figure 1. Younger participants gave a significantly lower security rating to the single-factor version (A) than older participants. Younger participants also gave a significantly higher rating to the two-factor version of the service (B) than older participants, who saw much less difference between the two services in terms of security.

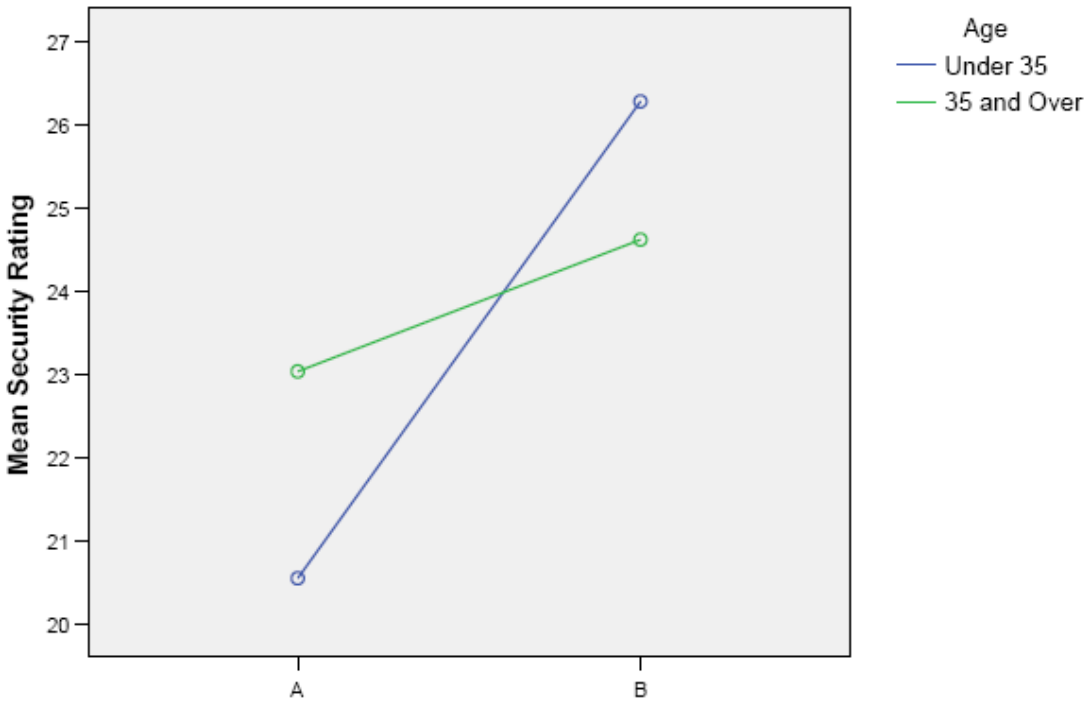

Figure 1: Security Rating by Age Group

As with the security item on the usability questionnaire, both age groups rated the two-factor version as being more secure than the single-factor version. However, according to the ratings scale, the younger age group judged this difference more salient than the older age group. The usability data did not show this difference in attitude towards the issue of security between the two age groups. However, the usability questionnaire gathers immediate reactions to the experience of using the service. The ratings scales are administered after experience of both versions are completed and may encompass consideration of other real-world factors such as the safe-keeping of the device itself and how the device operates.

Finally, in terms of *ease of use*, there was a highly significant difference between the two authentication methods, $p = 0.006$, with the single-factor version being seen as significantly more easy to use than the two-factor version. There were no significant interactions and no between-subjects effects. Again participants were asked to comment on their reasons for their response. One participant stated that "[*the single-factor] was easier because I am used to this kind of thing. The other one [two-factor version] would be difficult to remember carrying the device with me.*"

4.4 Rank Order Preference Results

The quality ratings obtained for each version were recorded simultaneously, allowing direct comparison of both versions. Therefore for each participant, the expressed quality score was also converted into a rank order for each version. The results are shown in Table 4. (Note again that one participant was unable to complete the debriefing interview due to time constraints.)

Ranked Best	Overall Preference	Convenience	Security	Ease of Use
Single-factor	32 (52.5%)	42 (68.9%)	4 (6.6%)	30 (49.2%)
2-factor	26 (42.6%)	9 (14.8%)	46 (75.4%)	9 (14.8%)
Rated equally	3 (4.9%)	10 (16.4%)	11 (18.0%)	22 (36.1%)

Table 4: Frequency Ranked Best

Examining the 58 participants whose quality ratings expressed an *overall preference* between the two authentication methods, a binomial test found no significant difference in votes for the two services; although slightly more participants expressed a preference for the single-factor version this was not significant.

In terms of *convenience*, excluding those who rated both services equally, significantly more participants rated the current, single-factor authentication method best (binomial, $p < 0.001$). The same was also true for *ease of use*, albeit a slightly less significant result (binomial, $p = 0.001$). In terms of *security*, however, significantly more participants rated the two-factor service variant best (binomial test excluding those who ranked both services equally, $p < 0.001$). Note that all four participants who ranked the single-factor service best in terms of *security* were in the older age group.

4.5 Relationship between Usability and Comparative Ratings

It was of interest to investigate possible relationships between usability perceptions (taken from the repeated measure attitude questionnaires) and the comparative ratings for quality, convenience, security and ease of use. Pearson correlations were computed between the differences in usability of the two authentication methods and the differences in each of the comparative ratings, as shown in Table 5.

Possible Relationship	Correlation
Usability and Overall Quality	0.46, 0.21, <.001
Usability and Convenience	0.42, 0.18, .001
Usability and Security	0.244, 0.06, .058
Usability and Ease of Use	0.54, 0.30, <.001
Notes: Displayed Data: Pearsons' r, R^2, p.	

Table 5: Correlations between Measures of Usability and Quality, Convenience, Security and Ease of Use for the Two Versions

The usability difference between the two versions was highly significantly correlated to the perceived difference in overall quality, convenience and ease of use between the two services. In each case a strong positive relationship was indicated.

The relationship between the difference in usability and the difference in security between versions, although approaching significance, $p = 0.058$, was not significant.

4.6 Qualitative Results

Towards the end of their research session, participants made comments on their likes, dislikes and preferences on a wide range of issues related to methods of authentication. Some typical responses are presented here.

4.6.1 Likes and Dislikes

The single-factor method was reported as being easy to use, short and straightforward. The main dislikes commented on for this version of the service related to the automated service in general rather than the method of authentication.

The two-factor method was reported for the additional security participants judged that it offered. The main dislikes and suggestions made in reference to this method included comments about disliking having to carry a device with them in order to use the service, that the procedure takes longer having more numbers to enter and that the numbers on the screen of the device could be made larger. It was suggested that the service should add a question at the beginning of the service "*Do you have your access device?*" in order to check that the user can proceed with this method of verification. A number of comments were also made suggesting users were unsure as to how the device operated or the means by which it provided additional security e.g. "*It doesn't really add extra security. There was no verification it was me.*"

4.6.2 Physical Attributes of Security Device

Participants were asked what they thought about the size of the hardware keyfob device they used: 80.4% thought it was "about the right size", with the remaining participants equally split between those who thought it was too big and those who thought it was too small. All participants stated that they thought the display was easy to read.

4.6.3 When to Use Security Device

Participants were asked when they thought they should be asked to use the security device in an automated telephone banking service: 54.1% stated that it should be in every call, whereas 31.1% stated that it should be used as enhanced security only in order to authorise a high value transaction; 13.1% of participants stated that they should not be asked use the device at all in this context.

4.6.4 Observations by Researchers

The researchers who guided participants through the research sessions recorded some observations on participants' handling of the device. Most, they noted, used their thumb to press the button to display the code. Almost all used only one hand to interact with the device. When inputting the code many placed the device on the desk but some kept it in their hand whilst typing the code, clamping the telephone handset under their chin using their neck and shoulder.

5 Discussion of Hypotheses

5.1 Hypothesis H1: Perceived Usability

There was evidence to support the hypothesis that the telephone banking services with single-factor and two-factor authentication approaches were significantly different in terms of perceived usability. The single-factor authentication process was found to be significantly more usable than the two-factor process, $p = 0.024$.

The single-factor version rated a mean score of 5.56, indicating that it was seen as a highly user-friendly method. The two-factor modified version of the telephone banking service was rated significantly lower with a mean score of 5.31. However, the usability score is still above 5.0 on the 7-point scale indicating it was a usable method in the broad sense.

In terms of the individual usability attributes, the single-factor approach was seen as superior to the two-factor approach in being less *confusing*, requiring less *concentration*, causing less *frustration*, being less *complicated*, not being *too fast*, being *easier to use* and not having too many *details to input*. In contrast, the two-factor approach was rated significantly higher on the attribute of *security*, and scored highly positively in this regard, although perceptions of security were reasonable for the single-factor process as well.

5.2 Hypothesis H2: Comparative Ratings

The comparative ratings included overall quality, convenience, security and ease of use.

In terms of *overall quality*, there was no evidence to support the hypothesis that the quality scores for the two different authentication approaches in telephone banking were significantly different. There were no significant interactions or between-subject effects.

In terms of *convenience*, there was significant evidence to support the hypothesis that the two different authentication approaches in telephone banking were perceived differently. The single-factor version was seen as significantly more convenient than the two-factor version.

In terms of *security*, there was also significant evidence to support the hypothesis that the two different authentication approaches in telephone banking were perceived differently. The two-factor version was seen as significantly more secure than the single-factor version. There was a significant interaction with age group. Younger participants appreciated the additional security provided by the two-factor approach, marking the single-factor version significantly lower than two-factor. However, on the ratings scale, older participants appeared to be less aware of the security issue, marking both services similarly, possibly due to a lack of understanding of how the technology works.

Finally, there was significant evidence to support the hypothesis that the *ease of use* scores for the two different authentication approaches in telephone banking were different. The single-factor version was seen as significantly easier to use than the two-factor version.

Mean usability scores were not significantly correlated with security ratings, but were significantly correlated with overall quality, convenience and ease of use. Higher security ratings were accompanied by lower usability, depicting the difficulty in providing usable security at the user interface.

5.3 Hypothesis H3: Preferences

There was no evidence to support the hypothesis that the two authentication methods were different in terms of overall preference. There were mixed opinions on the rank order, with 32 of the 61 participants preferring the single-factor version, and 26 preferring the two-factor version.

The single-factor version was ranked as the most *convenient* by a significant margin, and as *easier to use*; whereas, the two-factor version was ranked as the most *secure* by a significant margin.

6 Conclusions

This experiment explored user attitudes towards the usability and security of single-factor and two-factor methods for authentication in the context of an already established automated telephone banking service. The results show some interesting differences between the two authentication approaches which, together with participant preferences expressed in the interview, can inform decisions on the use of two-factor authentication involving physical tokens in a telephony context, and specifically a telephone banking service.

In terms of performance, unsurprisingly the two-factor process involving an additional stage took significantly longer to complete than the single-factor method ($p < 0.001$). Perhaps more surprisingly, use of the push-button device to obtain and enter a one-time access code caused relatively few problems for users, the large majority of whom (90%) successfully entered the access code at the first attempt on reaching this dialogue stage. The remainder did not respond or made errors when inputting the code in response to the initial prompt. However, most of these (five out of six) went on to successfully give the access code within the same call, with just one participant requiring an additional call to successfully navigate the access code stage. This is encouraging as it might have been expected that simultaneous use of the telephone handset and key fob under the (moderate)

time constraints imposed by the automated system prompts could cause significant difficulty for users.

In fact, data logged by the system showed that problems in the two-factor approach were actually caused by confusion between the secret number and the access code; both numbers being six digits in length, and used after one another. Nine participants (14.5%) failed at least one call in the two-factor version for this reason. Note, however, that the problem in this case was not due to interference in memory between multiple user 'passwords' in the sense described earlier, since in the experiment neither six-digit number had to be recalled from memory (the fixed secret number was supplied on the persona sheet given to each participant and the dynamic access code was displayed on the device as it would be in real life). It is possible that simultaneous display of both numbers (an artefact of the experiment) may have influenced participant behaviour in this regard; however it seems likely that this confusion is also likely to arise in real life where the access code at least is on display whilst participants are on the telephone during the automated dialogue. For this reason, it may be beneficial to avoid such similarities in the information required of the user. There may be less confusion if the one-time access code contains a different number of digits to the secret number.

Both approaches to customer authentication obtained a mean usability rating above 5.0 on the 7-point scale, indicating a generally positive reaction to the services (5.56 for the single-factor version and 5.31 for the two-factor version). There was, however, significant evidence that the single-factor knowledge-based authentication process was more usable than the two-factor process employing an additional one-time access code from a small, portable device. The single-factor approach was rated (moderately) significantly higher overall (p=0.024), and for seven of the twenty-two usability attributes measured in the experiment. The two-factor version, in contrast, scored significantly higher only on the issue of *security*.

Overall ratings for *ease of use, convenience* and *security* exhibited a similar pattern, with the single-factor approach considered significantly *easier to use* and more *convenient* than the two-factor version, but less *secure*. (Note again, however, that in general both versions were rated highly on each of these dimensions.)

Interestingly, however, the *overall quality* ratings for the two services were not statistically significantly different (both were rated highly). In terms of rank ordering also, although slightly more participants expressed a preference for the single-factor version this was not significant. Analysis of user comments showed that although participants liked the single-factor version as it was easy, straightforward and quick, they also valued the extra security provided by the two-factor service.

From a practical point of view these data are likely to be a comfort to those considering use of a similar (two-factor token-based) approach to authentication. Although the gain in security obtained through such an approach is associated with a decrease in the usability of the service it could be argued that the effect is relatively moderate, given users' (lack of) overall preference and the positive usability score awarded to the two-factor version. Thus, it may be that on balance the gain in security afforded by use of a token is of such magnitude and significance that this outweighs the apparently moderate cost in terms of usability. This, however, is a business discussion that is beyond the scope of this paper.

From a research point of view the results show that participants in the experiment did understand the added security provided by use of the physical token. As noted above however, this gain in security was obtained at a cost to the usability of the service. Results from the study thus corroborate the findings of previous research on two-factor authentication for Internet banking (Weir et al., 2009), which found that usability and convenience were sacrificed when added layers of security were added to the system. More generally, these data provide valuable *empirical evidence* of the trade-off between usability and security in automated systems that is not often evident in other research in this area. Crucially, moreover, it does so in an automated telephony context.

6.1 Future Work

This study indicated that an extended authentication procedure for automated telephone banking does boost perceptions of security, but at a cost to usability. However, any negative impact on usability may be guarded against by design changes after observations made during this study. As described above, one issue found was confusion between the secret number and the access code. More clarity could be achieved for users by differentiating the number of digits in these numbers. Alternatively, the access device could generate an alphanumeric code, which would further differentiate it from the secret number. The use of alphanumerics in an automated telephone dialogue would, however, entail an increase in complexity for the speech recognition system. It would be beneficial to measure any effects on usability and perceptions of security with these new authentication versions, together with any error rates or differences in call timings between the new versions.

An interesting facet of the results, in part related to this issue, is the suggestion that attitudes towards security may be affected by demographic factors such as the age of the user. Participants in the older age group rated the usability of the two-factor method significantly lower than their younger counterparts, at the same time rating the single-factor version higher. (Younger users rated the usability of the two versions much more similarly.) On closer inspection it was found that a majority of the participants who appeared to confuse the secret number and access code (seven out of nine) were in the older age group. Although based on small numbers of participants this may provide a clue to the difference in attitude between the two age groups with respect to usability. It may also be resolved in part by better differentiation of the secret number and access code as described above. Those in the older age group, however, also viewed the difference in security between the two authentication methods as being much less than younger participants, possibly due to a lack of understanding of the technology. Such results highlight the importance of including a wide range of end users in research on security in automated systems, and provide interesting pointers for future possible research focusing on demographic factors.

The hardware device itself was criticised as being inconvenient to carry around. It was also found that for some participants, the access device was a mystery in terms of what actual security it provided and how it worked. In a real world context, it may be possible to mitigate this with effective customer literature. In an experimental setting, this could be achieved by using draft customer communications as part of the research blueprint that the participant could read prior to interacting with a two-factor authentication service.

In this study, usability scores were higher for the version of the service experienced second. In a real life context, it would be assumed that users become more adept at using the service with experience of using it. In an experimental setting, accustomed usage could be accounted for by allowing more than one use with each version of the service. Any habituation effects could then be measured by administering the usability questionnaires after each usage as well as after each version.

Finally, this experiment examined use of the device in a particular setting, that of a relatively quiet environment with a desk and landline telephone. Researchers observed that participants on the whole either placed the device on the desk during the telephone call, or held the device in their hand whilst balancing the telephone under their chin using their neck and shoulder. It would be interesting, therefore, to examine the effects of different environments on their use of the device and the resulting usability and performance e.g. in the absence of a surface on which to place the device (as might occur when travelling or sitting on an easy chair) and/or using a mobile telephone, which is more difficult to fix under the chin in the manner described.

Appendix A. Items in Usability Questionnaire

Statements were presented in a randomised order for each participant.

Q1 I thought the service was too complicated.

Q2 When I was using the service I always knew what I was expected to do.

Q3 I thought the service was efficient.

Q4 I liked the voice.

Q5 I would be happy to use the service again.

Q6 I found the service confusing to use.

Q7 The service was friendly.

Q8 I felt under stress when using the service.

Q9 I felt this service was secure.

Q10 The service was too fast for me.

Q11 I thought the service was polite.

Q12 I found the service frustrating to use.

Q13 I enjoyed using the service.

Q14 I felt flustered when using the service.

Q15 I think the service needs a lot of improvement.

Q16 I had to enter too many details during the service.

Q17 I felt the service was easy to use.

Q18 I would prefer to talk to a human being.

Q19 I thought the voice was very clear.

Q20 I felt that the service was reliable.

Q21 I had to concentrate hard to use the service.

Q22 I did not feel in control when using the service.

References

Adams, A. and Sasse, M. (1999). “Users are not the enemy.” Communications of the ACM 42 (12), pp.41-46, reprinted (2005) in: Cranor and Garfinkel (Eds), *Security and Usability*, O’Reilly, pp.639-649 [chapter 32].

Bishop M. (2005). “Psychological acceptability revisited.” In: Cranor and Garfinkel (Eds), *Security and Usability*, O’Reilly, pp.1-11 [chapter 1].

Brostoff, S. and Sasse. M.A. (2000). “Are Passfaces more usable than passwords? A field trial investigation.” In McDonald S et al (Eds): 'People and Computers XIV - Usability or Else', Proceedings of HCI, Sunderland, UK, pp.405-424, Springer.

Chiasson, S., Forget, A., Stobert, E., van Oorschot, P.C. and Biddle, R. (2009). “Multiple Password Interference in Text Passwords and Click-Based Graphical Passwords.” Proceedings of the 16th ACM conference on Computer and Communications Security, Chicago, Illinois, USA, pp.500-511.

Clarke, N., Karatzouni, S. and Furnell, S. (2009). “Flexible and transparent user authentication for mobile devices.” *Proceedings of the 24th IFIP TC 11 International Information Security Conference*, pp.1-12.

consumeraffairs.com (2005) “Consumers losing confidence in online commerce, banking.” News article, June 28, 2005, last accessed September 2010.

Coolican, H. (1990). *Research Methods and Statistics in Psychology*, Hodder & Stoughton, GB.

Cranor, L. and Garfinkel, S. (2005) *Security and Usability: Designing Secure Systems That People Can Use*. O’Reilly Media, Inc.

Dhamija, R. and Perrig, A. (2000). “Déjà vu: a user study using images for authentication.” In Proceedings of the Ninth Usenix Security Symposium, pp.45-48.

Dunphy, P., Nicholson, J. and Olivier, P. (2008). “Securing passfaces for description.” Proceedings of the Fourth Symposium on Usable Privacy and Security (SOUPS’08), Pittsburgh, Pennsylvania, USA, pp.24-35.

Dunphy, P. Heiner, A.P. and Asokan, N. (2010). “A closer look at recognition-based graphical passwords on mobile devices.” Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS’10), Redmond, WA USA, Article No.3.

FFIEC, (2005). “FFIEC guidelines on authentication in internet banking environments.” Available from: http://www.ffiec.gov/pdf/authentication_guidance.pdf. (Accessed 23.03.2010).

Forget, A., Chiasson, S., van Oorschot, P.C. and Biddle, R. (2008). “Improving text passwords through persuasion.” Proceedings of the Fourth Symposium on Usable Privacy and Security (SOUPS’08), Pittsburgh, Pennsylvania, USA, pp.1-12.

Forget, A., Chiasson, S. and Biddle, R. (2007). “Helping users create better passwords: is this the right approach?” Proceedings of the Third Symposium on Usable Privacy and Security (SOUPS’07), Pittsburgh, Pennsylvania, USA, pp.151-152.

Foster J.C., Dutton R.T., Jack M.A., Love S., Nairn I.A., Vergeynst N.A. & Stentiford F.W.M., (1993). “Intelligent dialogues in automated telephone services.” In Baber C. &

Noyes J.M. (Eds), *Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computer,* London, Taylor and Francis, pp.167-175.

Florêncio, D. and Herley, C. (2007). "A large-scale study of web password habits." Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada, pp.657-666, ACM Press.

Gaw, S. and Felten, E.W. (2006). "Password management strategies for online accounts." Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS'06), Pittsburgh, Pennsylvania, USA, pp.44-45.

Hiltgen, A., Kramp, T. and Weigold, T. (2006). "Secure internet banking authentication." *IEEE Security and Privacy*, vol 4(2), March, pp. 21-9.

ISO International Standardisation Organisation (1991). "ISO/IEC 9126: Information technology – software product evaluation – quality characteristics and guidelines for their use."

ISO International Standards Organisation (1998). "ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) Part II: guidance on usability."

Ion, I., Langheinrich, M., Kumaraguru, P. and Čapkun, S. (2010). "Influence of user perception, security needs, and social factors on device pairing method choices." Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS'10), Redmond, WA USA, Article No.6.

Kainda, R., Flechais, I. and Roscoe, A.W. (2010). "Security and usability: analysis and evaluation." Proceedings Fifth International Conference on Availability, Reliability and Security (ARES 2010), Krakow, Poland, pp.275-282.

Karat, J., (1988). "Software Evaluation Methodologies." In Helander, M. (Ed.), *Handbook of Human Computer Interaction*, Amsterdam: North-Holland, pp. 891-903.

Knight, W. (2008). "The price of love." *Infosecurity, 5* (1):30-3.

Kuo, C., Romanosky, S. and Cranor, L.F. (2006). "Human selection of mnemonic phrase-based passwords." Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS'06), Pittsburgh, Pennsylvania, USA, pp.67-78.

Landauer, T.K., (1988). "Research methods in Human Computer Interaction." In Helander, M. (Ed.), *Handbook of Human Computer Interaction*, Amsterdam: North-Holland, pp. 905-928.

Likert, R. (1932). "A technique for the measurement of attitudes." *Archives of Psychology*, *140*.

Love S., Dutton R.T., Foster J.C., Jack M.A., Nairn I.A., Vergeynst N.A. and Stentiford F.W.M. (1992). "Towards a usability measure for automated telephone services." *Proceedings of Institute of Acoustics Speech and Hearing Workshop*, vol.14, no.6, pp.553-559.

Love S., Dutton R.T., Foster J.C., Jack M.A. and Stentiford F.W.M. (1994). "Identifying salient usability attributes for automated telephone services." *Proceedings of International Conference on Spoken Language Processing*, pp.1307-1310.

Love, S. (1997). *The Role of Individual Differences in Dialogue Engineering for Automated Telephone Services*. University of Edinburgh, PhD thesis.

Mannan, M. and van Oorschot, P.C. (2007). "Security and usability: the gap in real-world online banking." In Proceedings New Security Paradigms Workshop (NSPW'07), New Hampshire, USA, pp.1-14.

Murdoch, S., Drimer, S., Anderson, R. and Bond, M. (2010). "Chip and PIN is Broken". 2010 IEEE Symposium on Security and Privacy. doi: 10.1109/SP.2010.33.

Nielsen, J. (1993). Usability Engineering, Academic Press, USA.

O'Gorman, L. (2003). "Comparing passwords, tokens and biometrics for authentication." Proceedings of the IEEE, vol. 91(12), December, pp.2021-2040.

O'Gorman, L., Bagga, A. and Bentley, J. (2005). "Query-directed passwords." *Computers and Security*, Vol.24-7, October, pp.546-560.

O'Gorman, L., Brotman, L. and Sammon, M. (2006a). "How to speak an authentication secret securely from an eavesdropper." 14th Int. Workshop on Security Protocols, Cambridge, UK, March 2006.

O'Gorman, L., Brotman, L. and Sammon, M. (2006b). "Comparing authentication protocols for securely accessing systems by voice." 2nd Secure Knowledge Management Conference, Polytechnic University, Brooklyn, New York City. Sept. 2006.

Piazzalunga, U., Savaneschi, P. and Coffetti, P. (2005). "The usability of security devices." In Cranor,L.F. and Garfinkel, S. (Eds.), *Security and Usability: Designing Secure Systems that People Can Use*, O'Reilly Media, pp.221-42.

Preece, J., Rogers, Y. and Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons Inc., NY.

Robson, C. (1983). *Experiment, Design and Statistics in Psychology: An Introduction*, Pelican Books, GB.

Sasse, M.A., Brostoff, S. and Weirich, D. (2001). "Transforming the 'weakest link': a human-computer interaction approach to usable and effective security", BT Technology Journal 19(3), pp.122-31.

Sasse, M.A. (2004). "Usability and trust in information systems." In: *Trust and Crime in Information Societies*, edited by Robin Mansell and Brian S Collins, pp.319-348. Edward Elgar. ISBN 1 84542 177 9.

Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N. and Cranor, L.F. (2010). "Encountering stronger password requirements: user attitudes and behaviours." Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS'10), Redmond, WA USA, Article No.2.

silicon.com (2005). "Banks must boost security to drive online banking". Forrester Research New article, March 29, 2005, www.silicon.com, last accessed September 2010.

Sinclair, S. and Smith, S.W. (2005). “The TIPPI Point: towards trustworthy interfaces.” *IEEE Security and Privacy*, pp 68-71.

Toledano, D.T., Fernández Pozo, R., Hernández Trapote, A and Hernández Gómez, L. (2006). “Usability evaluation of multi-modal biometric verification systems”. *Interacting with Computers*, vol 18(5), September, pp.1101-1122.

Weir, C.S., Douglas, G., Carruthers, M. and Jack, M., (2009). “User perceptions of security, convenience and usability for eBanking authentication tokens.” *Computers and Security 28 (2009),* pp.47-62.

Weir, C.S., Douglas, G., Richardson, T. and Jack, M., (2010). “Usable security: user preferences for authentication methods in eBanking and the effects of experience.” *Computers and Security 22 (2010)*, pp.153-164.

Whiteside, J., Bennett, J. and Holtzblatt, K., (1988). “Usability engineering: our experience and evolution.” In Helander, M. (Ed.), *Handbook of Human Computer Interaction*, Amsterdam: North Holland, pp.791-817.

Wiedenbeck, S., Waters, J., Birget, J-C., Brodskiy, A. and Memon, N. (2005). “Authentication using graphical passwords: effects of tolerance and image choice.” Proceedings of the First Symposium on Usable Privacy and Security (SOUPS’05), Pittsburgh, Pennsylvania, USA, pp.1-12.

Yan, J., Blackwell, A., Anderson, R. and Grant, A. (2004). “Password memorability and security: empirical results.”, *IEEE Security and Privacy*, Vol. 2, Issue 5, pp.25-31.